Discrete stochastic optimization with continuous auxiliary variables

Rodolphe Le Riche*,
Alexis Lasseigne**, François-Xavier Irisarri**

*CNRS and Ecole des Mines de St-Etienne, Fr.
** ONERA, Composite Systems and Materials Dept., Fr.

From multi-level to auxiliary variables (1)

Most engineering and physical systems are described at different parameterization levels.

Example : heat consumption , f , in cities

entire city
(coarser level)

districts
v
cost , efficiency tradeoffs

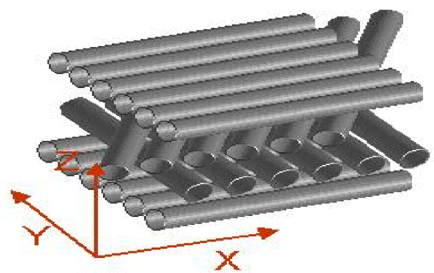
buildings
(finer level)
x
which heater & insulation chosen from a catalog

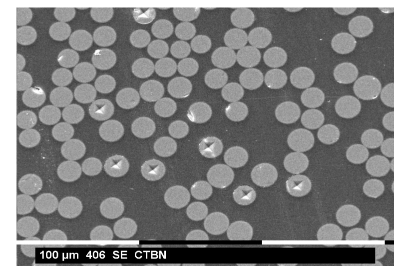
scale change

From multi-level to auxiliary variables (2)

Most engineering and physical systems are described at different parameterization levels.

Example : composite structure performance , f





full structure
(coarser level)

meso-scale
v
local stiffness

micro-scale
(finer level)
x
fiber position

scale change

From multi-level to auxiliary variables (3)

We want to optimize the system (here a composite structure)

$$\min_{x \in S} f\left(y\left(x\right)\right) \quad \text{or, in short,} \quad \min_{x \in S} f\left(x\right)$$

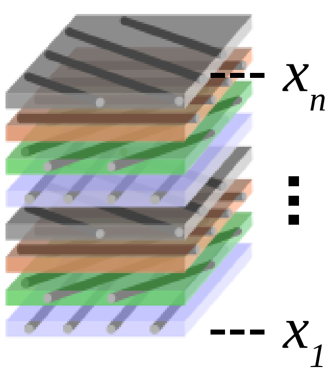
y(x) numerical simulator of the structure (stress, strains, strength, mass, ...) , **numerically expensive**

v(x) is numerically inexpensive.

H1

v (auxiliary variables, plate stiffnesses)

$x \xrightarrow{v(x)} v$



x (fiber positions in plies, discrete because of manufacturing tools and regulations)

x(v) usually doesn't exist, therefore *f(v)* neither.
E.g., there are many fiber positions for one choice of plate stiffnesses.

From multi-level to auxiliary variables (4)

Often, x is discrete and
v is seen (i.e., mathematically approximated) as continuous.

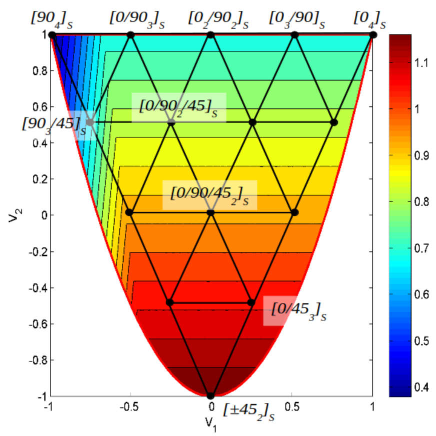
H2

(not necessary but used here)

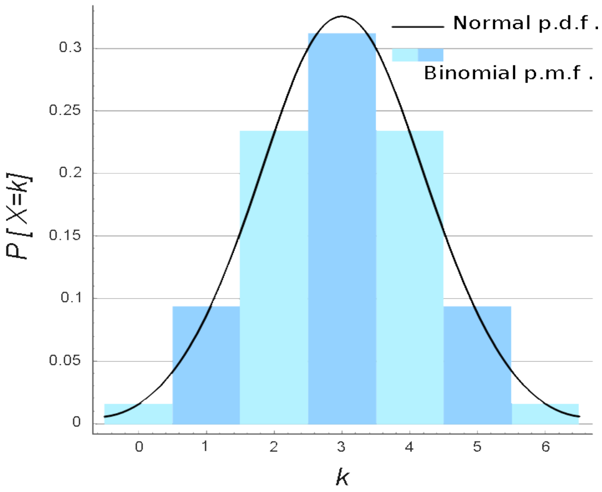
The continuous approximation is more accurate as $n = dim(x)$ increases

Composite material performance :

buckling as a function of v_3 and v_4



- $x \equiv$ fiber positions, discrete because of manufacturing
- $v \equiv$ local plate stiffness or lamination parameters

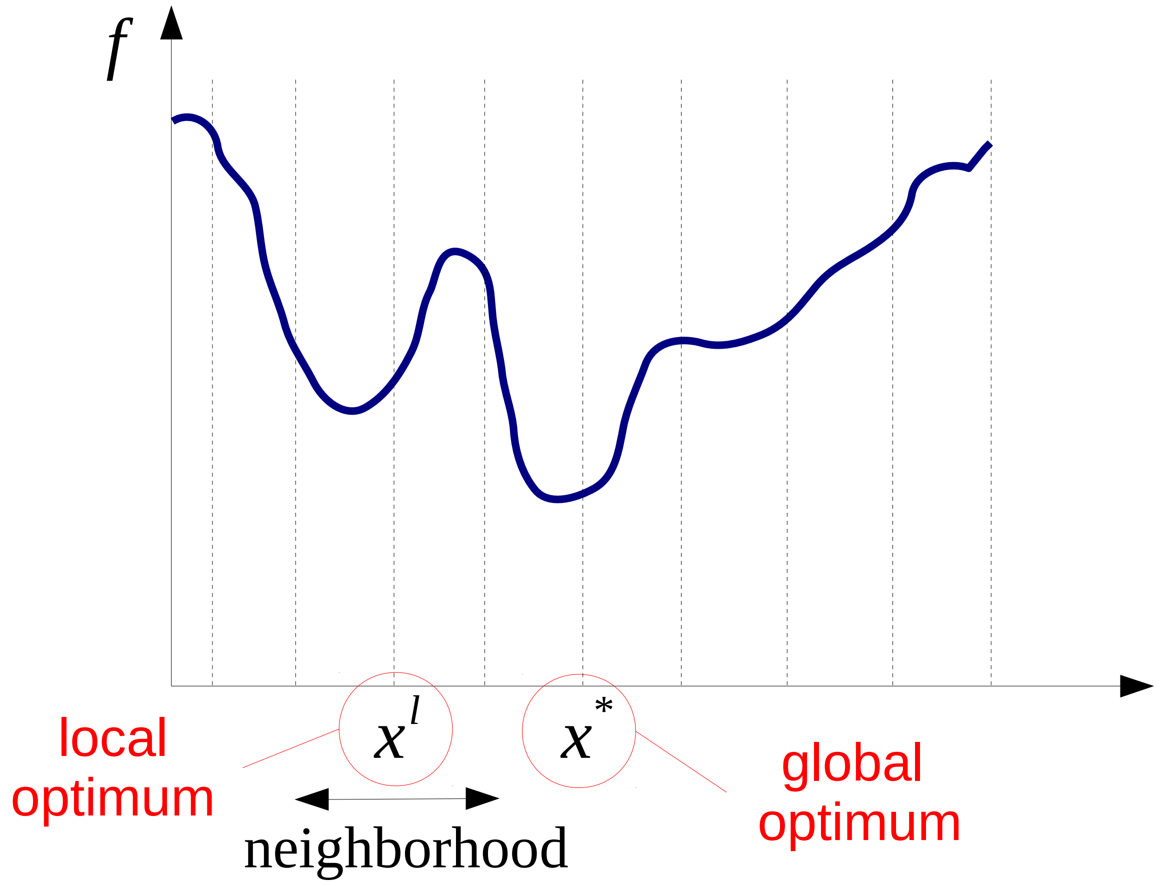


AN : statisticians do such approximations
Binomial law, n sufficiently large,
$\sim N(np, np(1\text{-}p))$

Goal : discrete global optimization

$$\min_{x \in S \subset \mathbb{N}^n} f(x)$$

but most of what will be said could be generalized to continuous and mixed optimization $S \subset \mathbb{R}^n$ or $\{\mathbb{R}^{n1} \cup \mathbb{N}^{n2}\}$



Flow chart of a general stochastic optimizer

- **Initialize** $p^{(t)}(x)$ sampling dist. for candidate points
  λ number of samples per iteration
  t time counter (nb. calls to f)

- **Sample** λ **candidates** $\{x^{t+1}, \dots, x^{t+\lambda}\} \sim p^{(t)}(x)$

- **Calculate their performances** $f(x^{t+1}), \dots, f(x^{t+\lambda})$

- **Learn the distribution**
  $p^{(t+\lambda)}(x) = Update(\ x^1, f(x^1), \dots, x^{t+\lambda}, f(x^{t+\lambda})\ )$
  or more often
  $p^{(t+\lambda)}(x) = Update(\ p^t(x), x^{t+1}, f(x^{t+1}), \dots, x^{t+\lambda}, f(x^{t+\lambda})\ )$

- **Stop or [** $t = t+\lambda$ **and go back to Sample]**

with different p's if x is continuous or discrete or mixed.

Discrete variables : The Univariate Marginal Density Algorithm (UMDA)

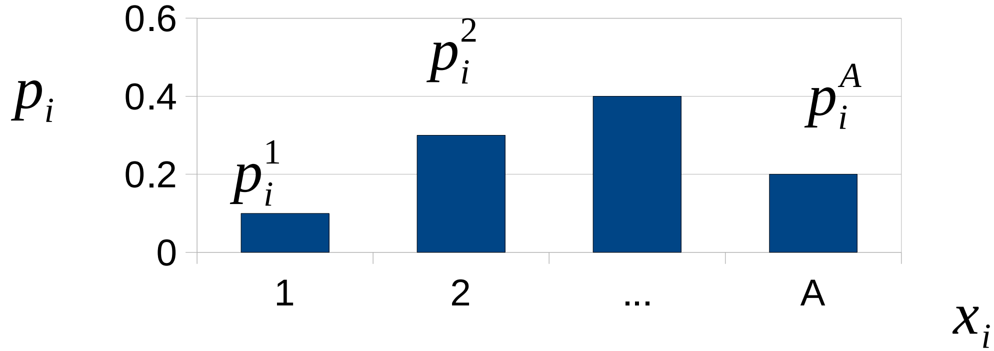
(Baluja 1994 – as PBIL – and Mühlenbein 1996)

$x \in S \equiv \{1,2,\ldots,A\}^n$ (alphabet of cardinality A)
e.g. $\{-45^o , 0^o , 45^o , 90^o\}^n$ (fiber orientations)
e.g. $\{\text{matl1} , \ldots , \text{matlA}\}^n$ (material choice)

The algorithm is that of a population based stochastic optimization (see before).

$p^{(t)}$ assumes variables independence (+ drop ^{t}), $\quad p(x) = \prod_{i=1}^{n} p_i(x_i)$

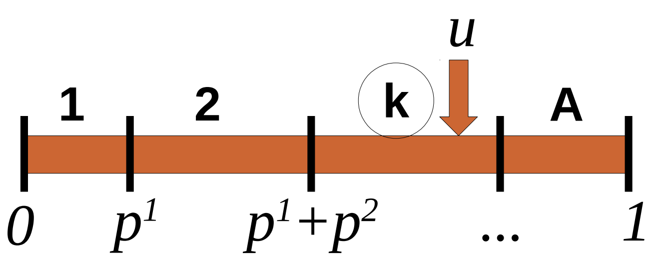
The x_i's follow an A-classes categorical law



$$\sum_{j=1}^{A} p_i^j = 1$$

UMDA (2)

Sampling



For $i=1,n$

spin a roulette wheel ,

$x_i = k$, k designated by $u_i \sim U[0,1]$

according to the p_i^j's

Learning

Select the μ best points out of λ ,

p_i^j is the frequency of j value at variable i in the μ bests

Guarantee $p_i^j \geq \varepsilon$ for ergodicity and $\sum_{j=1}^{A} p_i^j = 1$

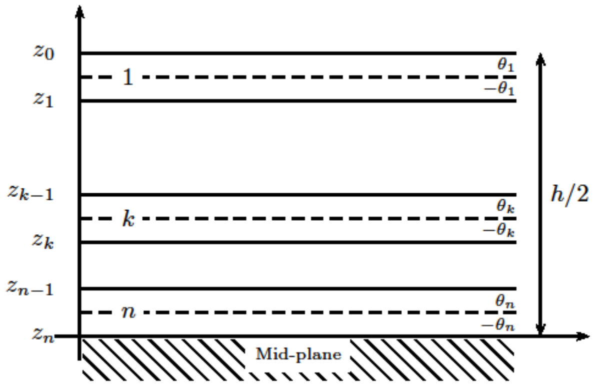
Application to composite design for frequency (1)

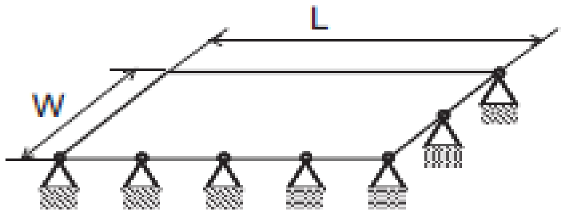
(from Grosset, L., Le Riche, R. and Haftka, R.T., A double-distribution statistical algorithm for composite laminate optimization, SMO, 2006)

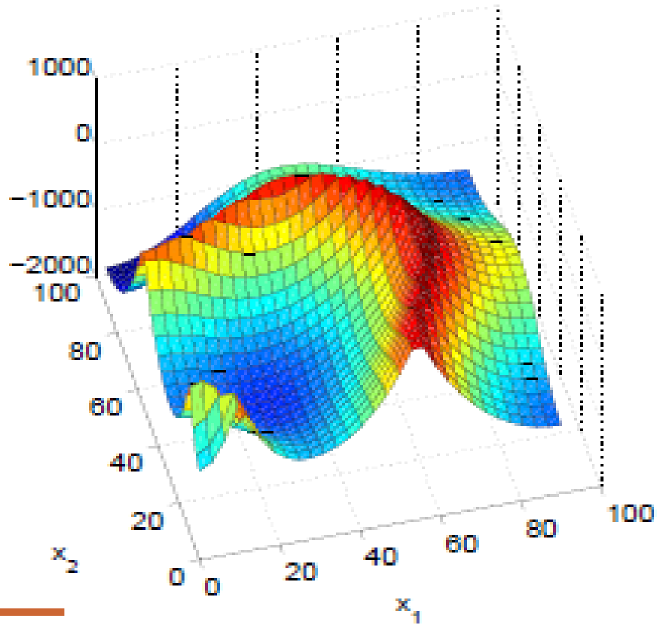
$max_x\, f_1(x_1,\ldots,x_{15})$, the first eigenfreq. of a simply supported plate

such that $0.48 \le \nu_{eff}(x) \le 0.52$

where $x_i \in \{0^o, 15^o, \ldots, 90^o\}$



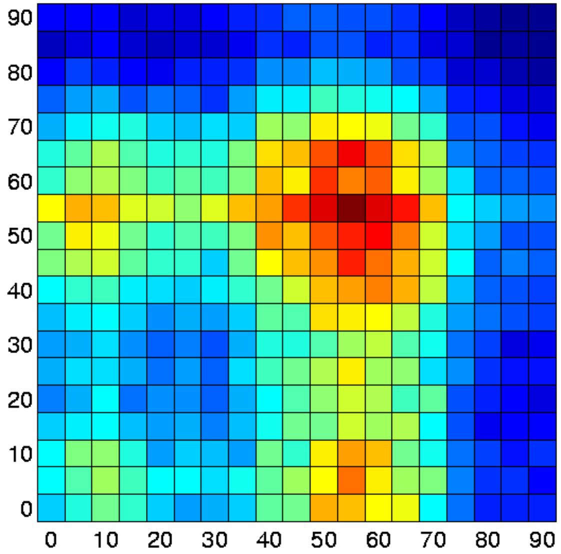




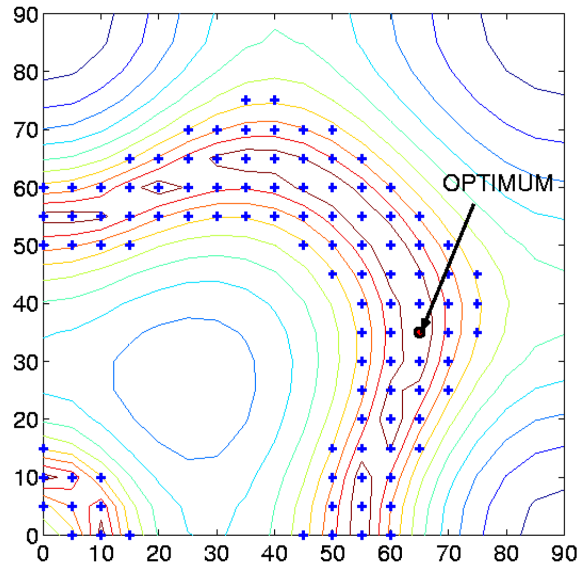
the constraint is enforced by penalty and creates a narrow ridge in the design space

Application to composite design for frequency (2)

density learned by UMDA (2D)



contour lines of the penalized objective function

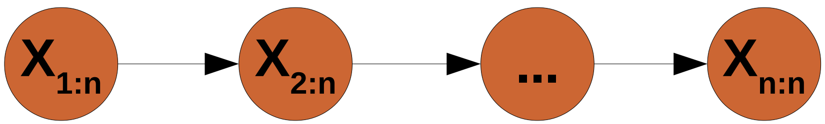


Independent densities can neither represent curvatures nor variables' couplings.

(from Grosset, L., Le Riche, R. and Haftka, R.T., A double-distribution statistical algorithm for composite laminate optimization, SMO, 2006)

Stochastic discrete optimization : learning the variables dependencies

More sophisticated discrete optimization methods attempt to learn the couplings between variables. For example, with pairwise dependencies :



$$p(x) = p(x_{1:n})\, p(x_{2:n}|x_{1:n}) \ldots p(x_{n:n}|x_{n-1:n})$$

Trade-off : richer probabilistic structures better capture the objective function landscape but they also have more parameters → need more *f* evaluations to be learned.

MIMIC (Mutual Information Maximizing Input Clustering) algorithm : De Bonnet, Isbell and Viola, 1997.
BMDA (Bivariate Marginal Distribution Algorithm) : Pelikan and Muehlenbein, 1999.

Multi-level parameter optimization with DDOA

(from Grosset, L., Le Riche, R. and Haftka, R.T., A double-distribution statistical algorithm for composite laminate optimization, SMO, 2006)

Mathematical motivation : create couplings between variables using two distributions (1 independent in x).

Numerical motivation : take into account expert knowledge in the optimization to improve efficiency.
E.g. in composites, the lamination parameters v (the plate stiffnesses) make physical sense.

Example in composites
Use of the lamination parameters

- v = lamination parameters = geometric contribution of the plies to the stiffness.

in-plane $$v_{\{1,2\}}(x) = \frac{1}{n}\sum_{i=1}^{n}\left\{\cos(2x_i), \cos(4x_i)\right\}$$

out-of-plane or flexural $$v_{\{3,4\}}(x) = \frac{1}{n^3}\sum_{i=1}^{n}\left((n-i+1)^3-(n-i)^3\right)\left\{\cos(2x_i), \cos(4x_i)\right\}$$

- Inexpensive to calculate from x (fiber angles).

- Simplifications : fewer v's than fiber angles. Often, the v's are taken as continuous.

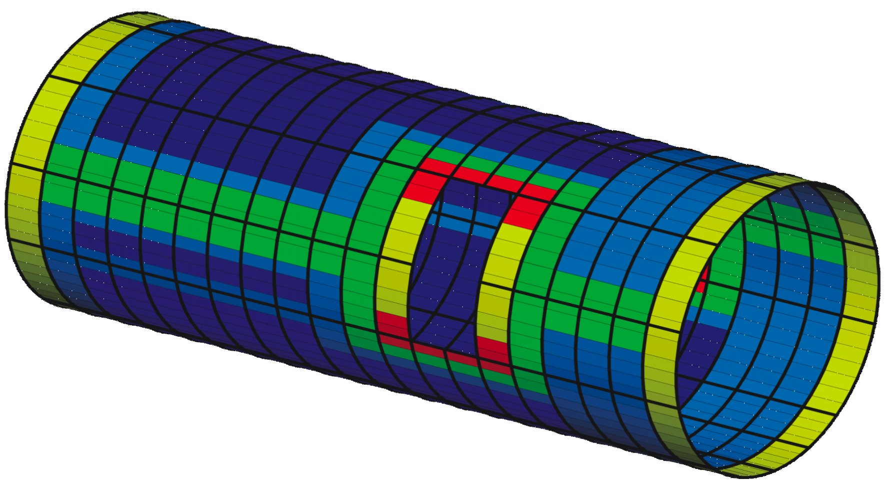
- But $f(v)$ typically does not exist (e.g., ply failure criterion).

Related past work in composites

(Liu, Haftka, and Akgün, « Two-level composite wing structural optimization using response surfaces », 2000.
Merval, Samuelides and Grihon, « Lagrange-Kuhn-Tucker coordination for multilevel optimization of aeronautical structures », 2008.)

Initial problem :

Optimize a composite structure made of several assembled panels by changing each ply orientation
→ many discrete variables



Decomposed problem :

Structure level
Optimize a composite structure made of several assembled panels by changing the lamination parameters of each panel
→ few continuous variables

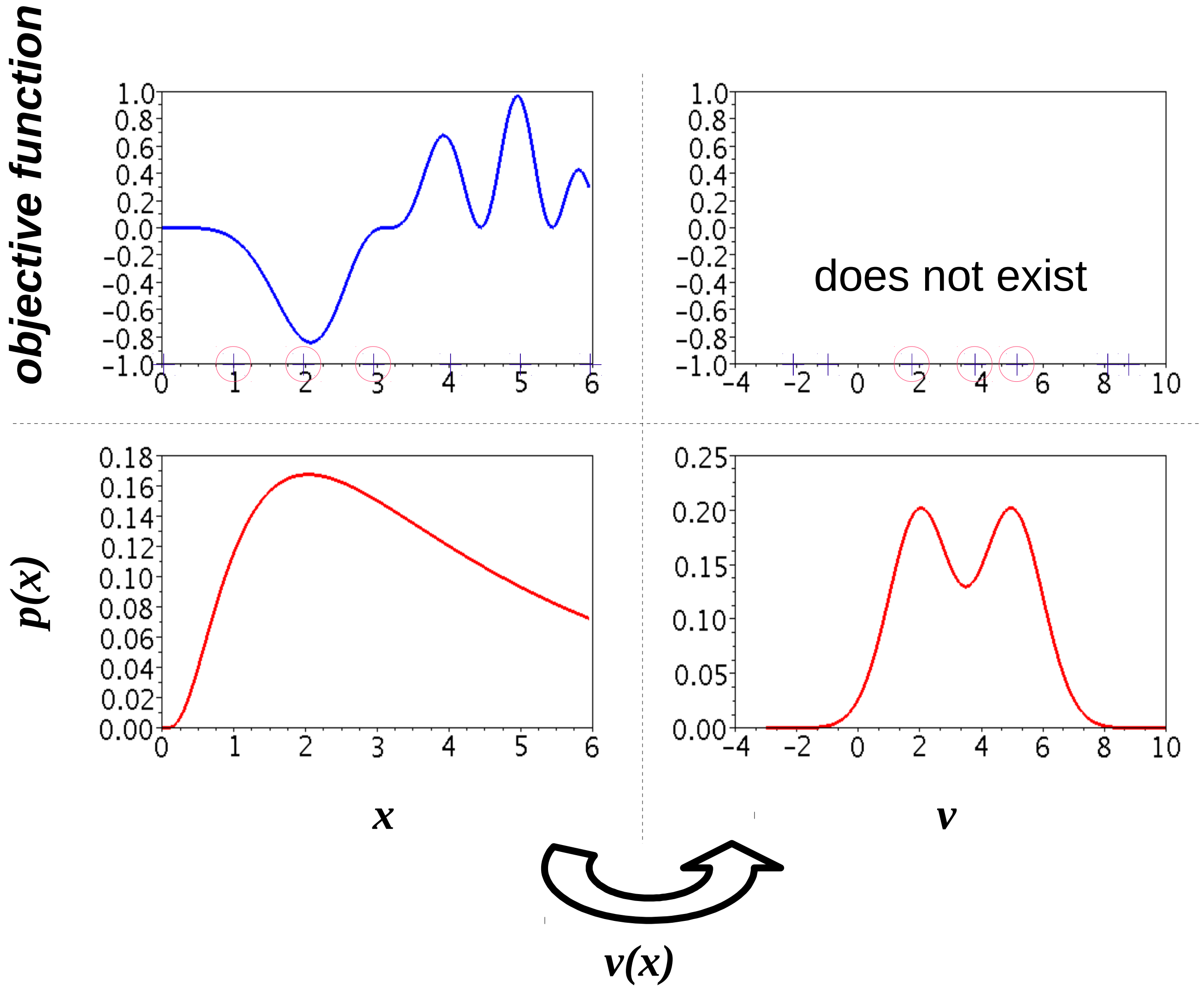
↓ **optimal *v*'s**

Laminate level
Minimize the distance to target lamination parameters by changing the ply orientations
→ few discrete variables

BUT for such a sequential approach to make sense, $\hat{f}(v)$ must exist and guide to optimal regions (i.e., prohibits emergence of solutions at finer scales).

The DDOA stochastic optimization algorithm

v(x) is costless → learn densities in the *x* AND *v* spaces at the same time.

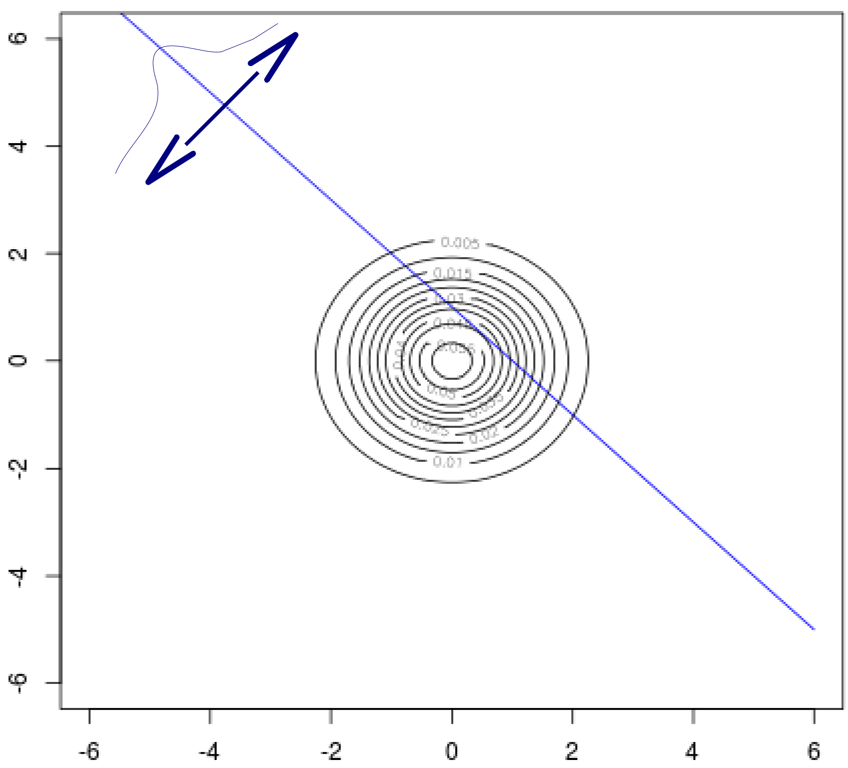


$$p_{\mathrm{DDOA}}(x) = p_{X \mid v(X)=V}(x) = p_{X \mid v(X)=v}(x) \times p_V(v)$$

The DDOA algorithm : $X \mid v(X)=V$?

Simple mathematical illustration : $v = x_1 + x_2$

$$p_X(x) = \frac{1}{2\pi}\exp\left(\frac{-1}{2}x^T x\right) \qquad p_V(v) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}(v-1)^2\right)$$



Intermediate step for a given v :

$$p_{X \mid v(X)=v=1}(x)$$

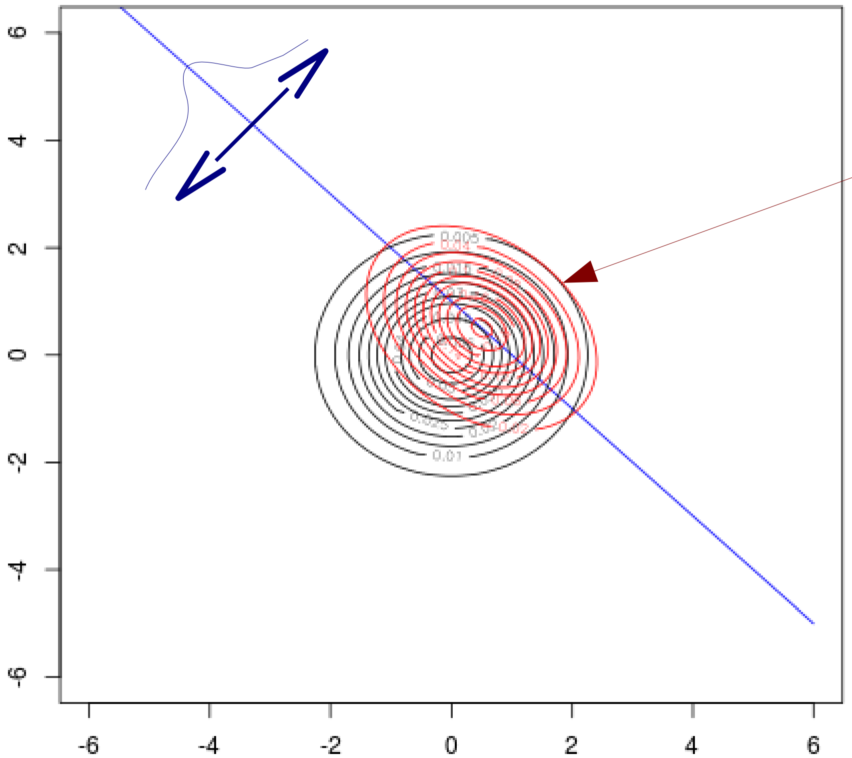
is a degenerated Gaussian along

$$x_1 + x_2 = 1$$

(cross-section of the 2D bell curve along the blue line + normalization)

The DDOA algorithm : $X \mid v(X)=V$?

$$p_{X|v(X)=V}(x) = p_{X|v(X)=v}\, p_V(v(x)) = \dots = \frac{1}{2\pi} \exp\left(\frac{-1}{4} (x_1 - x_2)^2 - \frac{1}{2} (x_1 + x_2 - 1)^2 \right)$$



$p_{X|v(X)=V}(x)$

is a coupled distribution that merges the effects of X and V

Analytical calculation in the Gaussian case. In practice, use simulations ...

The DDOA algorithm (flow chart)

```
Choose λ, μ, ρ such that ρ>>1 and λ>μ
Initialize p_V(v)and p_X(x)

For i=1,λ  do

   Sample v^target from p_V(v)
   Sample ρ>>1  x's from p_X(x)
   x(i) = the closest x to v^target
   Calculate f(x(i))

end For

Rank   x(1:λ), … , x(λ:λ) the proposed points
Update p_V(v) and p_X(x) from x(1:λ), … , x(μ:λ)

Stop ?
If no, go back to top ...
```

sampling of X | v(x)=V

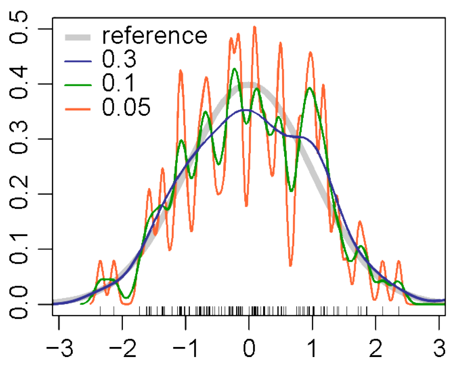
The DDOA algorithm (implementation)

p_X : cf. UMDA algorithm

p_V : isotropic gaussian kernel density estimation

$$p_V(v) = \frac{1}{\mu} \sum_{i=1}^{\mu} \frac{1}{(2\pi)^{d/2} \sigma^d} \exp\left(\frac{-1}{2\sigma^2} (v - v^i)^T (v - v^i) \right)$$

σ tuned by maximum likelihood



$p_V(v)$ for various σ's
(illustration from wikipedia)

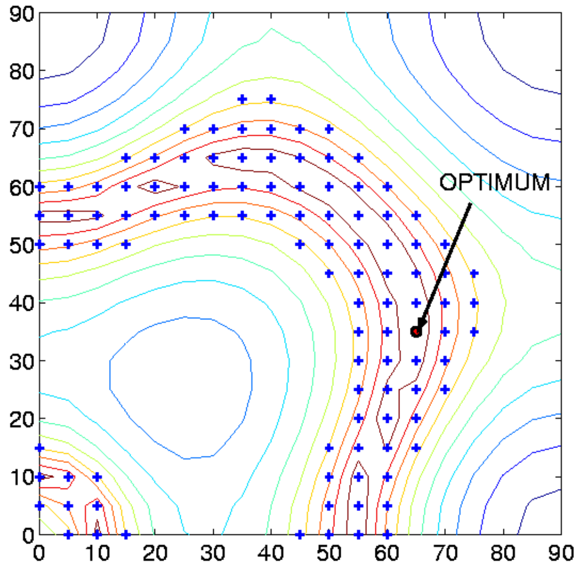
$d < n$,

$d = 2$ or 4 in composite applications

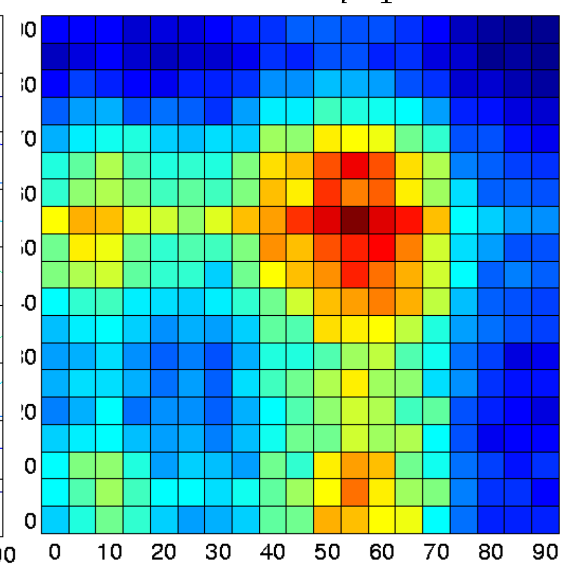
Application of DDOA to composite design for frequency

- $p_X(x)$ and $p_V(v)$ can be simple densities, without variables couplings (→ easy to learn), yet $p_{DDOA}(x)$ is a coupled density.

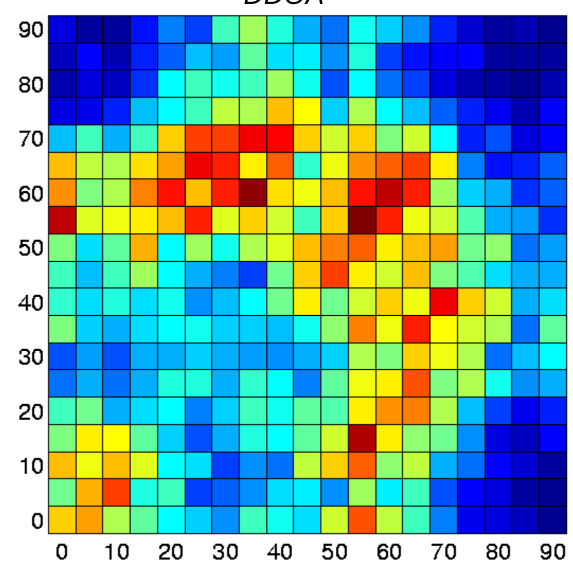
f(x) and selected points



$$p_X(x) = \prod_{i=1}^{n} p_i(x_i)$$



$p_{DDOA}(x)$



One half of the algorithm searches in a low dimension space.

Applications of DDOA to composite design (1)

A large number of tests were carried out in L. Grosset's PhD thesis (2004).
Multi-modal problems, A=5, x_i in $\{ 0 , 22.5 , 45 , 67.5 , 90 \}$

- Problems with coupled variables were created by enforcing constraints s.a.

  lower bound ≤ effective Poisson's ratio ≤ upper bound

  through penalty functions.

- **Extensional problem** :
  - max transverse stiffness such that low. bnd. ≤ Poisson's ratio ≤ upp. bnd.
  - all information in 2 in-plane lamination parameters v_1, v_2

- **Extensional-flexural problem** :
  - min CTE_x such that 1st eigenfrequency $\geq \omega_{min}$
  - all information capture by 4 lamination parameters v_1, v_2 (in-plane) v_3, v_4 (flexural)

- **Strength problem** :
  - max load at 1st ply failure (at any x_i)
  - not all information in the 2 in-plane lamination parameters v_1, v_2

and now a summary of results ...

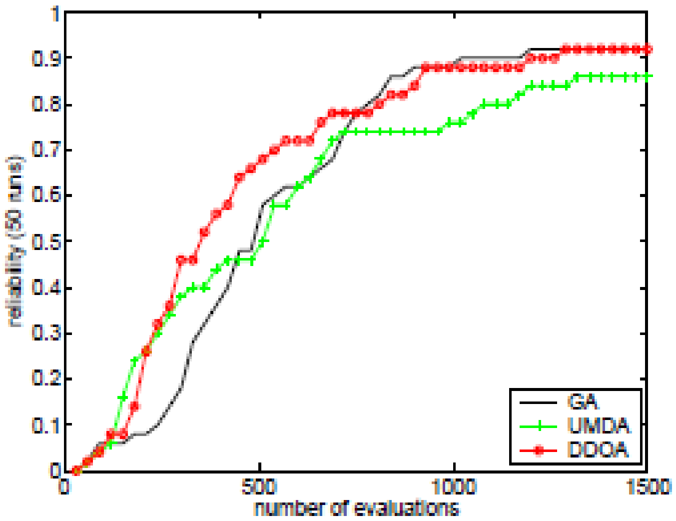
Applications of DDOA to composite design (2)

- Performance measure :
  reliability over 50 runs = probability of finding the optimum

- DDOA has an increasing advantage over a Genetic Algorithm (GA) and UMDA as the number of variables n increases.

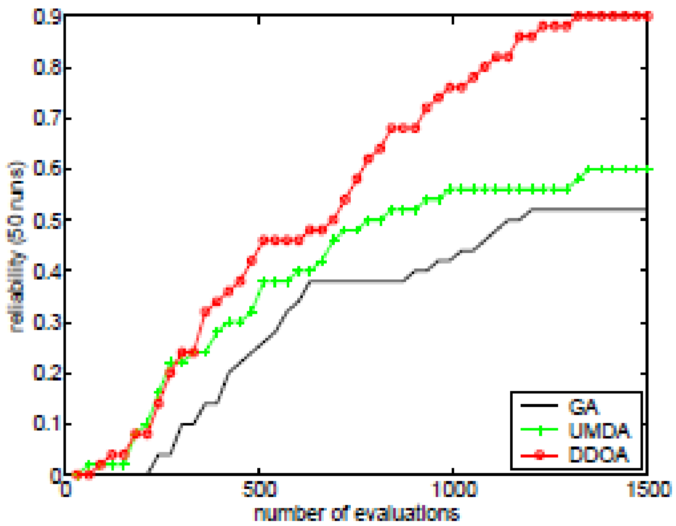
- Ex : extensional-flexural problem

$n = 6$



size(S) = $5^6 = 15625$

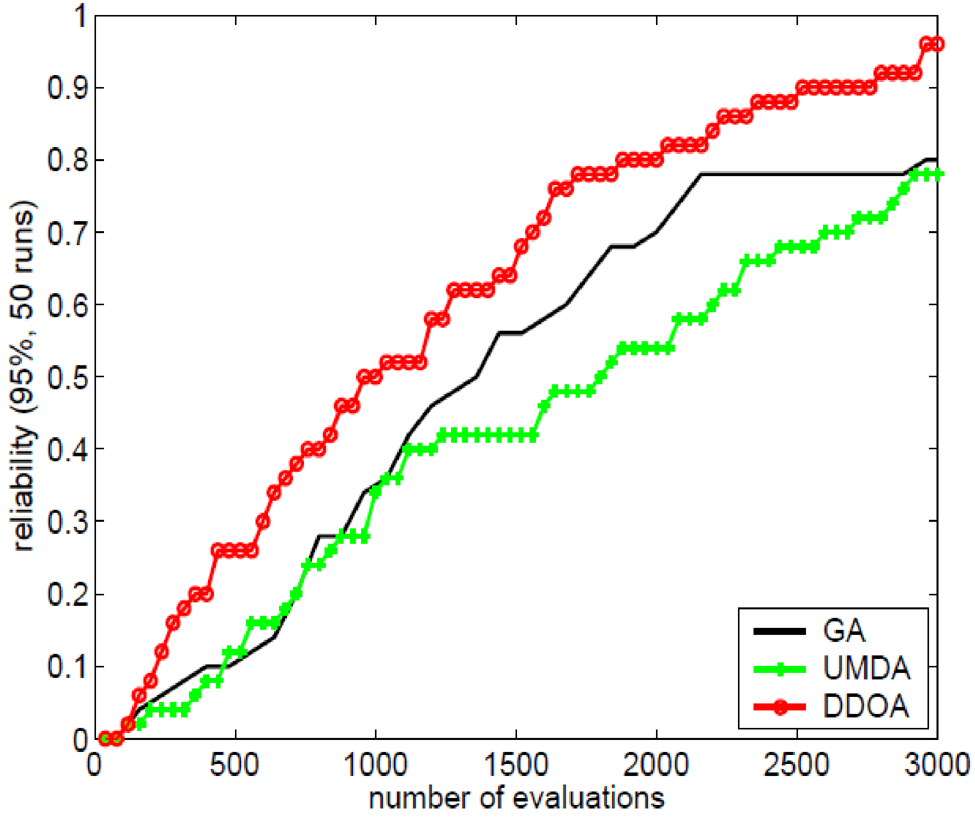
$n = 12$



size(S) = $5^{12} \approx 244$ millions

Applications of DDOA to composite design (3)

Comparison of 3 algorithms whose parameters have been optimized on the laminate strength problem



Applications of DDOA to composite design (4)

L. Grosset's work showed the potential of DDOA.

The sampling algorithm was not implementing the theory,

$$p_{DDOA}(x) = p_{X \mid v(X) = v}(x) \, . \, p_V(v)$$

sample $\nu \geq \lambda$ x^i's from $p_X(x)$
sample λ v^i's from $p_V(v)$
for i=1,λ do
    choose $x^i = arg\ min_i\ min_j \ ||\ v(x^i) - v^j\ ||$
    remove x^i and associated v^j from the sets
end

pair and discard

→ We are further investigating this algorithm with F.X. Irisarri and A. Lasseigne (on-going)

Double Densities for composite design (1)

Generalization of the use of Double Densities : DDEA

Like DDOA but instead of "sample $p_X(x)$", do

crossover + mutation + permutation of an x in the EA population

Optimization and comparison of many algorithms

UMDA, Evolutionary Algorithm (EA) specialized for composites (notably through a permutation operator), DDOA and DDEA (new).

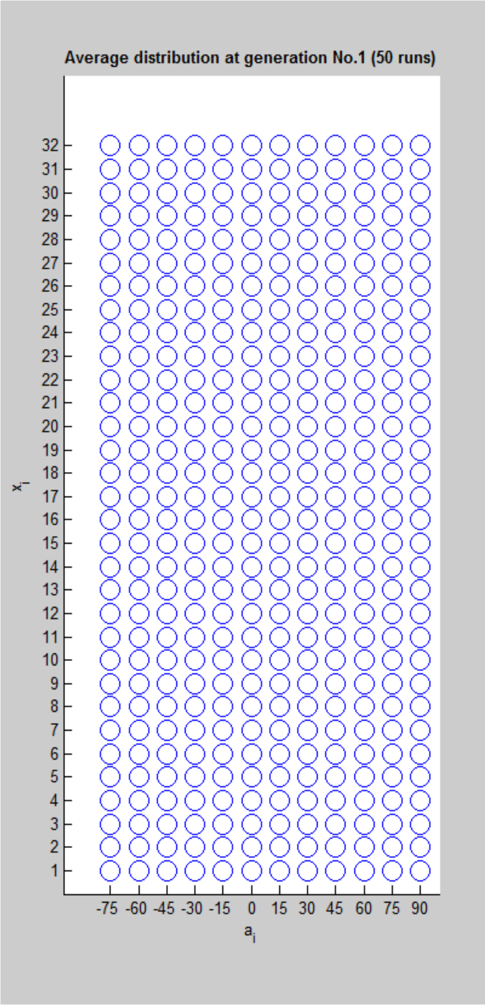
First tests on a buckling load maximization problem,

n=32 , A=12

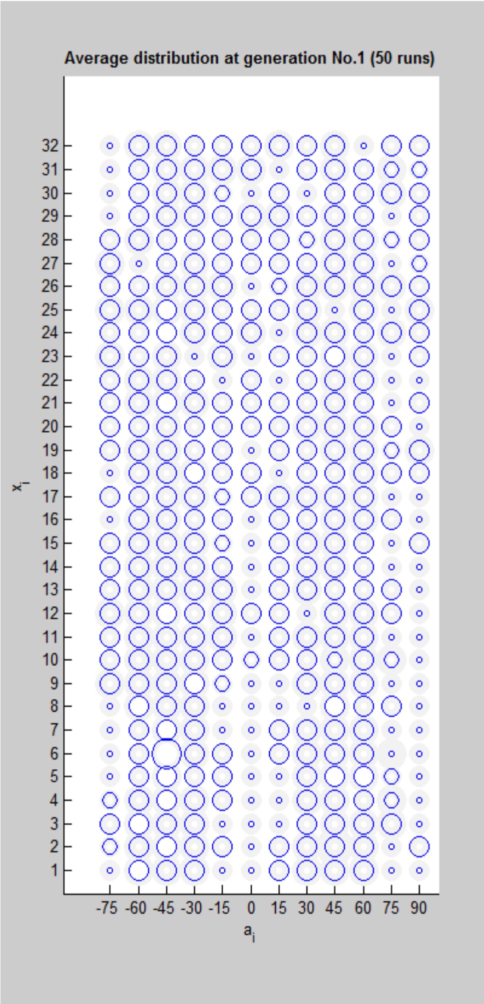
Double Densities for composite design (2)

frequencies of variables values, buckling problem, n=32 , A=12

UMDA

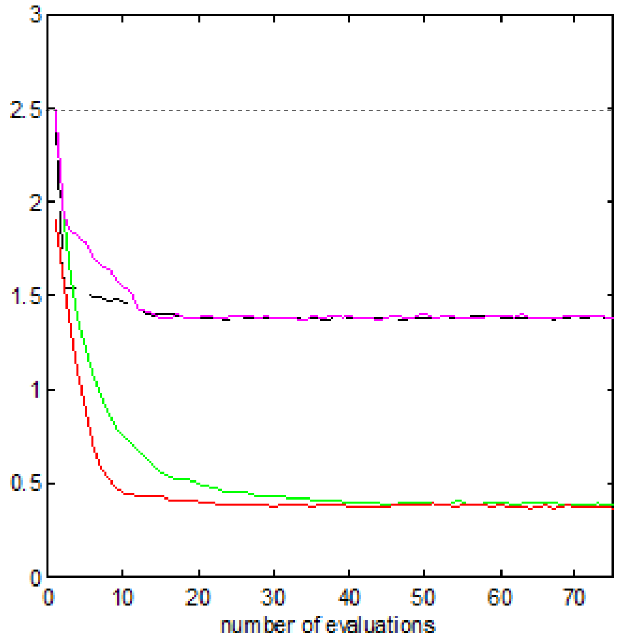


DDOA

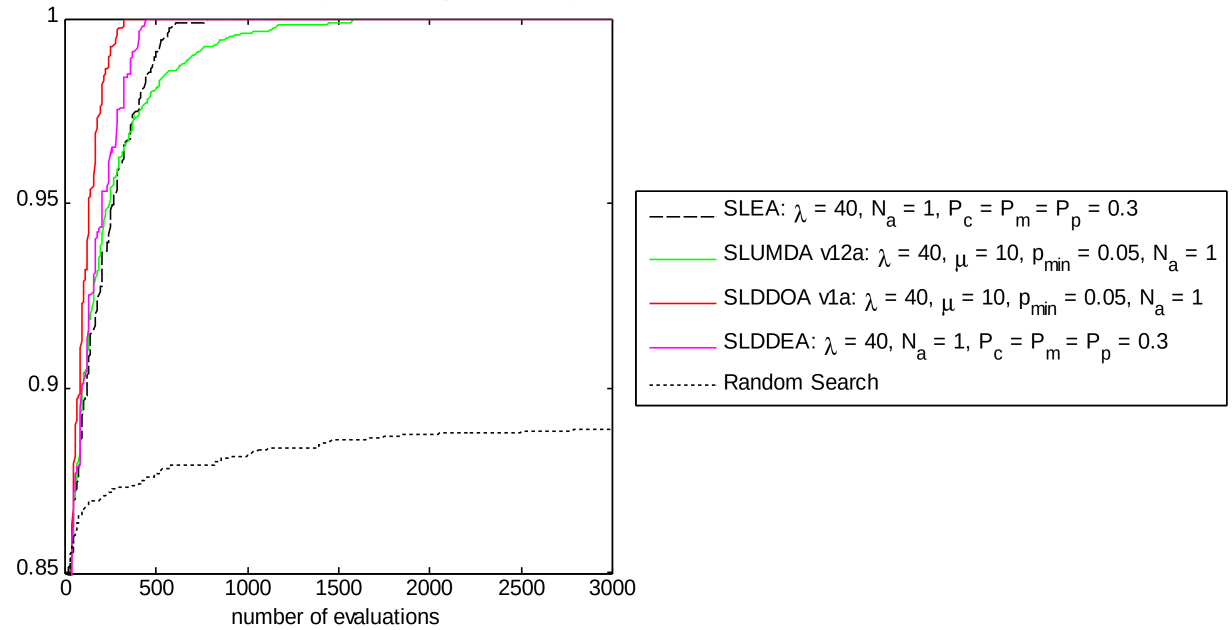


Double Densities for composite design (3)

median entropy
of $p_X^{(t)}$



median normalized buckling
load (1 is optimal)



Confirms the efficiency gain on the buckling problem for both EA and UMDA but this was not observed on in-plane problems (under investigation).

End of the episode ...